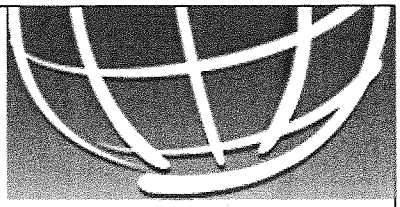
MAP4C: Foundations for College Mathematics, Grade 12, College Preparation

Unit 4: Data Management

Activity 4: The Line of Best Fit



Overview | Expectations | **Content** | Assignment

Content

Drawing the Line of Best Fit

A **line of best fit** describes how the points are distributed in a scatter plot. It is drawn through a set of data points in a scatter plot only when there is an apparent pattern in which the points fall. The line of best fit goes through as many of the data points as possible. If there does not appear to be a relationship (i.e. there is no correlation), no line should be drawn.

The Equation of the Line of Best Fit

The **equation of the line of best fit** is an algebraic summary of the relationship between two variables that appear to be linearly related.

Graphing calculators (e.g.TI-83 Plus) and graphing software (e.g. MS Excel) are capable of determining the equation of the line of best fit when a scatter plot displays a linear trend (i.e. the points seem to form a straight line).

Once a line of best fit is drawn, its equation can also be determined.

Remember that the **general equation for a straight line** is

$$y = mx + b$$

where y represents the dependent variable;
x represents the independent variable;
m represents the slope, also known as the rate at which y changes as x changes and;
b represents the y-intercept.

The **strength** of a correlation between the two variables is measured by the **correlation coefficient (*r*)**. The closer *r* is to the value of 1 or -1, the stronger the correlation. A positive sign represents a positive correlation while a negative sign represents negative correlation. The correlation coefficient is obtained from the coefficient of determination (r^2) by taking its square root.

For example, the scatter plot below (Figure 1) has a correlation coefficient, $r = 0.99$ ($\sqrt{0.98} = 0.99$, since we want r, not r^2). Since $r = 0.99$ is very close to 1, the scatter plot has a very strong positive correlation (the points on the scatter plot form an almost perfectly straight line).

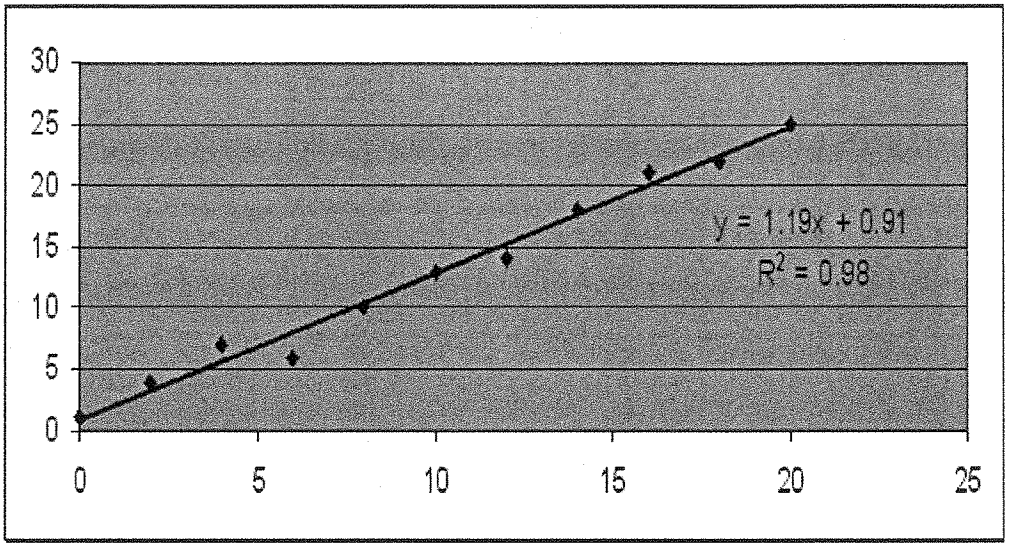


Figure 1

On the other hand, the scatter plot below (Figure 2) has a correlation coefficient, $r = -0.88$ ($\sqrt{0.78} = 0.88$, since we want r, not r^2, and the negative sign is for the negative direction of the scatter plot). Since $r = -0.88$ is not as close to -1 as -0.99, the scatter plot has a weaker and negative linear correlation (the points on the scatter plot are more spread apart).

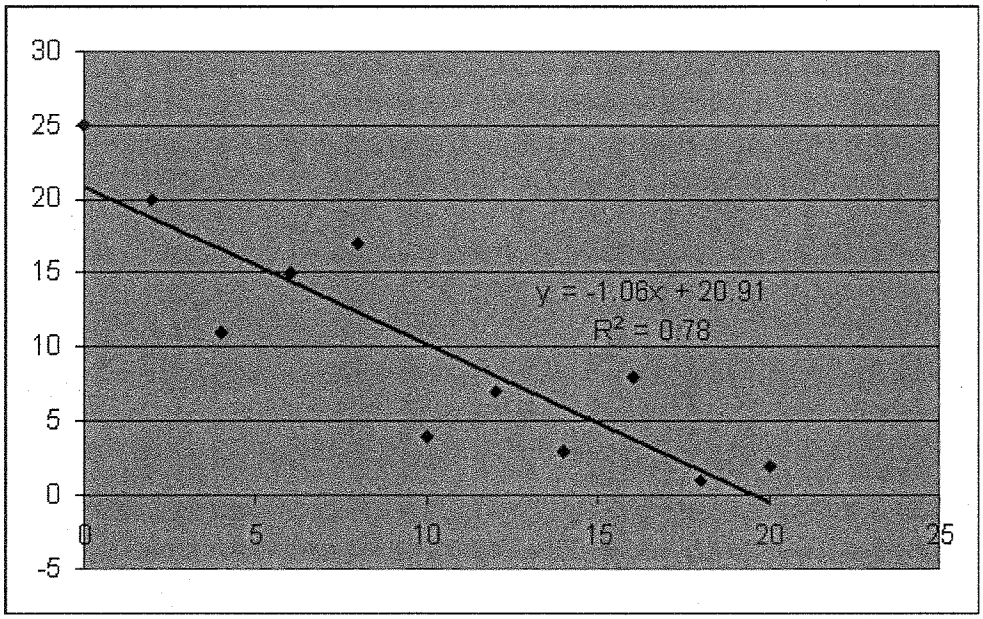
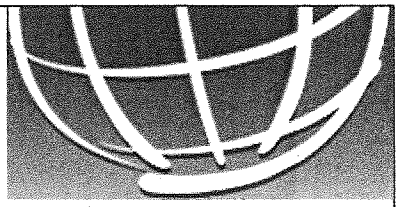


Figure 2

MAP4C: Foundations for College Mathematics, Grade 12, College Preparation

Unit 4: Data Management

Activity 4: The Line of Best Fit



Overview | Expectations | **Content** | Assignment

Content

e.g.

Liam collected data on a person's heart rate and the number of minutes spent riding a stationary bike at a constant speed.

Amount of Time Riding the Bike (Minutes)	Heart Rate (Beats per minute)
0	70
2	78
4	82
6	85
8	87
10	90
12	92
14	95
16	97
18	99
20	102

Use graphing technology to create a scatter plot and draw a line of best fit for the data.

Then, determine the equation of the line of best fit and comment on the strength of the correlation.

Solution

The scatter plot and the line of best fit for this set of data looks like this:

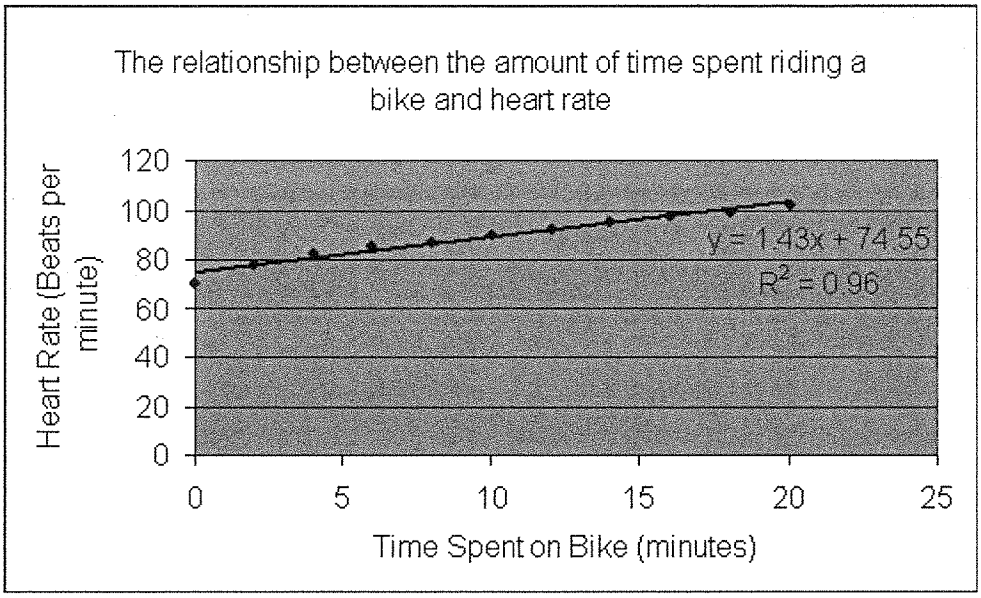


Figure 3

The equation of the line of best fit is

$y = 1.43x + 74.55$

where *y* represents the person's heart rate (beats per minute) and *x* represents the amount of time (minutes) spent riding the bike.where y represents the person's heart rate (beats per minute) and x represents the amount of time (minutes) spent riding the bike.

The strength of the correlation is $\sqrt{0.96} = 0.98$ (since we want ***r***, not r^2), which means that the correlation is a strong and positive one. This is confirmed by looking at the scatter plot.

Try This

Learn how to draw a line of best fit and determine its equation using **MS Excel**.

Learn how to draw a line of best fit and determine its equation using the **TI-83 Plus**.

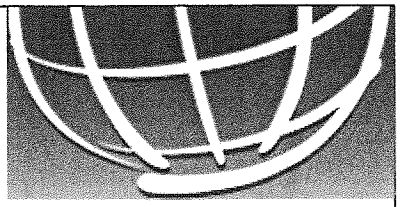
BACK | NEXT

This is a disclaimer. External Resources will open in a new window. Not responsible for external content.

MAP4C: Foundations for College Mathematics, Grade 12, College Preparation

Unit 4: Data Management

Activity 4: The Line of Best Fit



Overview | Expectations | **Content** | Assignment

Content

Using the Line of Best Fit to Make Predictions

The equation of the line of best fit is useful to make predictions to solve problems. **Interpolation** is the process of predicting values that fall *within* the range of data collected. **Extrapolation** is the process of predicting values that fall *outside* the range of data collected.

Example

Liam recorded the monthly average temperature and the number of ice-cream sales of his store. He created a scatter plot and noticed a linear relationship with the equation as shown below:

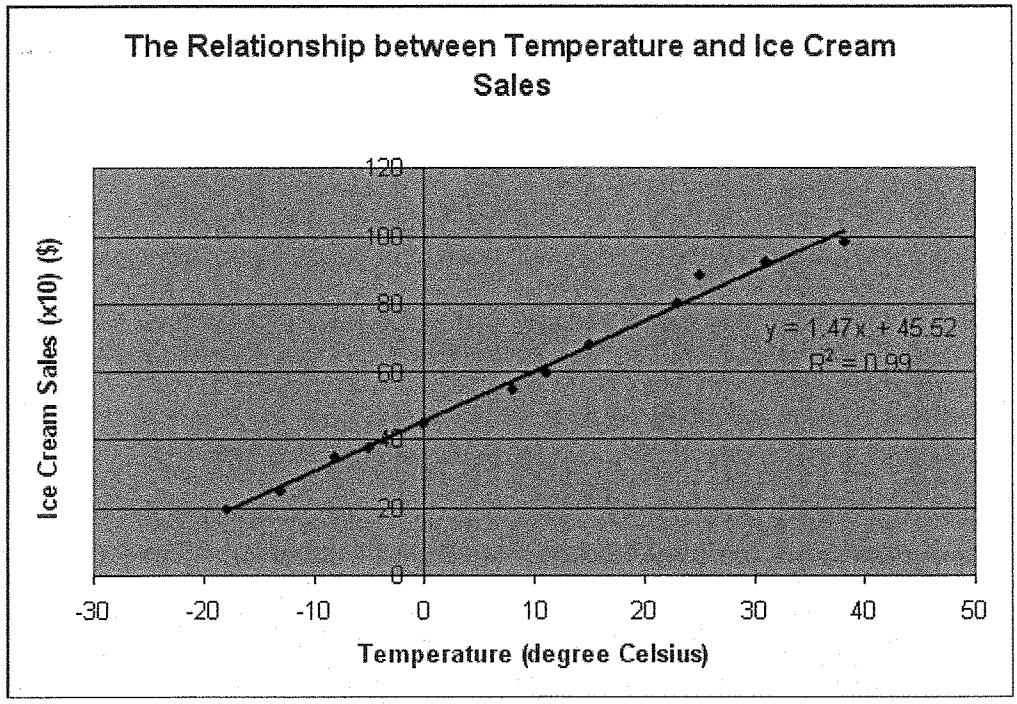


Figure 18

1. Use the equation of the line of best fit to predict:

a. The ice cream sales when the temperature is 25°C.
b. The temperature when the ice-cream sales is $1200. Comment on the reasonableness of this prediction.

2. Comment on the strength of the correlation.

Solution

1. The equation of the line of best fit is

$$y = 1.47x + 45.52$$

where *y* represents the ice cream sales (in 10's of dollars) and *x* represents the temperature (in degree Celsius).

a. To determine the ice-cream sales when the temperature is 25°C, we substitute $x = 25$ into the equation:

$$\begin{aligned} y &= 1.47(25) + 45.52 \\ &= 82.27 \end{aligned}$$

We have to multiply this value by 10 to get the actual sales amount. Therefore the ice-cream sale when the temperature is 25°C is $822.27.

Since the prediction falls within the range of the data (i.e. 25°C falls within the data collected for approximately -20°C to 35°C), this process is called *interpolation*.

b. To determine the temperature when the ice-cream sales is $1200, we substitute (we must divide by 10 since the scale is in tens) into the equation and solve for *x*:

$$\begin{aligned} 120 &= 1.47x + 45.52 \\ 120 - 45.52 &= 1.47x \\ 74.48 &= 1.47x \\ \frac{74.48}{1.47} &= x \\ x &= 50.67 \end{aligned}$$

Using the equation of the line of best fit, the predicted temperature would be 50.67°C.

Since the prediction falls outside the range of the data (i.e. 50.67°C falls outside the temperature range of -20°C to 35°C), this process is called *extrapolation*.

Using this equation to predict the temperature is not very reasonable, because firstly, a temperature of over 50°C is very high and the linear model assumes that ice cream sales will continue to increase as temperature increases. However, we know that temperature does not increase or decrease indefinitely – it usually levels off at a certain maximum or minimum temperature.

2. The value of the correlation coefficient is $r = \sqrt{0.99} = 0.99$ (since we want r, not r^2), which indicates a very strong, positive linear relation. We can verify this by looking at the scatter plot – the points form an almost perfectly straight line.

BACK | NEXT

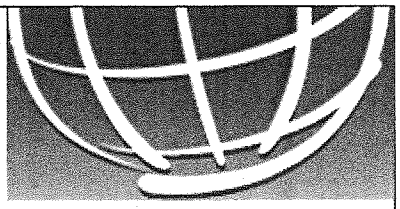
This is a disclaimer. External Resources will open in a new window. Not responsible for external content.

Overview | Expectations | **Content** | Assignment

MAP4C: Foundations for College Mathematics, Grade 12, College Preparation

Unit 4: Data Management

Activity 4: The Line of Best Fit



Overview | Expectations | **Content** | Assignment

Content

Practice

Questions

Data on the average rainfall during the growing season (in millimetres) and the average mass of corn grown (in grams) are collected as shown in the table below.

Amount of Rainfall (in mm)	Mass of Corn (in grams)
60	55
75	72
82	70
85	80
88	87
90	85
97	95
108	100
115	108
130	115

Use graphing technology, where applicable, to answer the following questions.

1. Create a scatter plot and determine the line of best fit.

   Answer

2. Determine the equation of the line of best fit and comment on the strength of the correlation.

   Answer

3. Predict the mass of the corn if 120mm of rain has fallen.

   Answer

4. Predict the amount of rainfall if the mass of the corn grown is 122 g.

   Answer

Solutions on website.

Resources

1. Unless otherwise indicated, all images in this Activity are from the public domain or are © clipart.com or Microsoft clipart and are used with permission.

BACK | NEXT

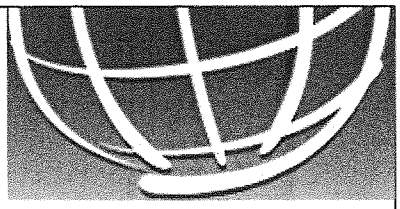
This is a disclaimer. External Resources will open in a new window. Not responsible for external content.

Overview | Expectations | **Content** | Assignment

MAP4C: Foundations for College Mathematics, Grade 12, College Preparation

Unit 4: Data Management

Activity 4: The Line of Best Fit



Assignment

Assignment

Complete this assignment and submit your work to the dropbox.

Before you upload your file, ensure your name appears on the top of every page of your document.

1. Use graphing technology, where applicable, to answer the following questions.

A neurologist wishes to determine if age affects how quickly a person reacts in a stressful situation, such as being asked on the spot to talk about themselves for a minute.

Data on the ages and reaction times of a group of randomly chosen people are collected and shown below.

Age (years)	Reaction Time (seconds)
10	8
12	7.5
12	8
15	7
19	6.5
22	6
25	5

26	4
29	4.5
32	3.5
32	3
35	3
38	2.8

a. Create a scatter plot and determine the line of best fit.
b. Determine the equation of the line of best fit and comment on the strength of the correlation.
c. Predict the reaction time of a 30 year old.
d. Predict the age of a person whose reaction time is 2 seconds. Comment on the reasonableness of this prediction. What other factors may influence a person's reaction time?

Resouces

1. Unless otherwise indicated, all images in this Activity are from the public domain or are © clipart.com or Microsoft clipart and are used with permission.

This is a disclaimer. External Resources will open in a new window. Not responsible for external content.